Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 8, No. 2, May, Pp. 563-570

Public opinion analysis of presidential candidate using naïve bayes method

Asno Azzawagama Firdaus*[1], Anton Yudhana[2], Imam Riadi[3]
Master Program of Informatics, Universitas Ahmad Dahlan, Indonesia[1]
Department of Electrical Engineering, Universitas Ahmad Dahlan, Indonesia[2]
Department of Information System, Universitas Ahmad Dahlan, Indonesia[3]

Article Info
Keywords:
Sentiment Analysis, Naïve Bayes, President, Python, Twitter

Article history:
Received: February 22, 2023
Accepted: April 18, 2023
Published: May 31, 2023

Cite:
A. A. Firdaus, A. . Yudhana, and I. . Riadi, "Public Opinion Analysis of Presidential Candidate Using Naïve Bayes Method ", KINETIK, vol. 8, no. 2, May 2023. https://doi.org/10.22219/kinetik.v8i2.1686

*Corresponding author.
Asno Azzawagama Firdaus
E-mail address:
2207048008@webmail.uad.ac.id

Abstract
Elections for president and vice president will take place in 2024. Heading into the election, promoted candidates were vying for public sympathy. People often discussed as presidential candidates are Anies Baswedan, Ganjar Pranowo, and Prabowo Subianto. Therefore, we need a way to predict potential candidates and voter demographics from public opinion on Twitter using sentiment analysis. One of his methods commonly used to classify sentiment analysis is Naive Bayes. This study used the naive Bayes classifier and the TF-IDF extraction function to add weights to the text. Use the scikit-learn Python library to help determine the polarity of negative and positive sentiment classes in your dataset. The datasets used were Twitter datasets acquired from October to December 2022, for a total of 15,000 datasets. The best test scenario obtained by splitting the test and training data is 70% test data and 30% training data, with the highest accuracy generated from the 95% Ganjar dataset. Using the Anies, Ganjar, and Prabowo test data, the positive mood scores for each candidate were 833, 77, and 524, respectively, while the negative mood scores were 637, 1423, and 976, respectively. The test was performed using a confusion matrix and k-fold cross-validation, and the best results were obtained on the Ganjar data set. That is a confusion matrix of 94.93% and a k-fold cross-validation of 94.46%. The lowest f1-score for the positive class is 67% for the Anies dataset and 27% for the negative class for the Ganjar dataset.

1. Introduction

General elections, or elections for short, are the activities carried out to elect leaders who occupy senior seats in the executive and legislative bodies of the Indonesian government. Elections are the embodiment of a country that adheres to democracy. An election process determines leaders from county/city, state to national level. Similarly, presidential elections have so far implemented systems in elections.

According to the General Election Commission (KPU), its website has announced the stages and schedule for the upcoming 2024 elections. President and Vice President nominations are from October 19, 2023, until November 25, 2023 [1]. Due to this fact, several polling agencies publish studies on presidential and vice presidential candidates. Based on one of the polling agency's websites, Poltracking Indonesia [2] has released its latest poll results for the 2024 presidential election. It indicated that the person's name is Ganjar Pranowo (Central Java Governor), Prabowo. Subianto (Chairman of the Gerindra Party) and Anies Baswedan (Governor of DKI Jakarta).

Statistical data shows smartphone users in 2021 were 6.3 billion [3]. This also encourages the use of elections for the dissemination of information through smartphone users. One of its applications is the use of social media as a campaign tool because it is considered a solution by the public in sharing their opinions freely. Among these social media, one of the most popular social media platforms is Twitter [4]. Twitter is one of the fastest-growing social networks, allowing users to interact with others anytime, anywhere, from their computer or mobile device. Twitter is a microblogging tool that allows users to freely express themselves [5]. Since its launch in July 2006, Twitter's user base has increased. As of September 2018, Twitter is estimated to have approximately 326 million registered users. Twitter is an effective forum that Indonesians, including individuals, businesses, and politicians, can use to disseminate information. Twitter allows users to send short messages (so-called tweets) of up to 140 characters. The tweet itself can consist of a text message and a photo [6].

Internet use has become a necessity for teenagers and adults [7]. This is encouraged by the development of knowledge and technology continues to increase so that new ideas and ideas are generated in their fields. Language identity is used to become aware of or apprehend the language of textual content so that it will expect the herbal language derived from written textual content [8]. One application of technology used in policymaking is sentiment analysis. Sentiment analysis is the computational study of moods, feelings, and opinions of texts. Sentiment analysis

aims to define user sentiment through sentences, documents, or other aspects of text [9]. Given a set of text documents containing beliefs about an object, opinion mining extracts the annotated object attributes and components of each document and determines whether the comments have positive or negative connotations. increase. Whether this is the case should be determined [10].

Classification methods are used to measure the success rate of classes. One classification method used to classify text is Naive Bayes. Naive Bayes has proven to be a well-known algorithm for performing sentiment analysis on text. Naive Bayes methods are supported in Python libraries to perform various domain classifications [11]. Naive Bayes is the approach of choice. This is because it represents a relatively accurate effect, is fast, and has proven to be an excellent tool for validating mood assessments [12]. If we want to improve the context of sentences in textual content classification systems, applying phrase collection, i.e., TF-IDF, also increases the educational process's accuracy level. M. Liang (2022) showed in his study [13] that using TF-IDF is a significant total value in weighting words within text categories.

V.A. Fitri et al. (2019) Discussed in their work [14] sentiment analysis on anti-LGBT issues using naïve Bayes methods, decision trees, and random forests. The study found that Naive Bayes was the most accurate classification method, with 86.43%. Furthermore, m. (2023) in their study [15] described mood analysis using a Twitter dataset on vaccination skepticism against Covid-19. This study combined Doc2Vec, Count Vectorizer, and TF-IDF to use Random Forest, Logistics Regression, Decision Trees, Linear SVC, and Naive Bayes methods. The best results in this investigation were obtained in experiments combining the TextBlob library, TF-IDF feature weighting, and the Linear SVC classification method with an accuracy of 0.96752. Then A. M. U. D. Khanday, et al (2022), in their research [16] using the extraction feature, revealed that TF-IDF could also optimize the classification process. The study took the topic of hate speech with a dataset on Twitter with an accuracy of 97%.

Discussions about the Indonesian President's candidacy on Twitter began to attract public sympathy in various circles. Tweets began popping up with the discussion. This is certainly a problem because it requires an approach that is able to filter all public opinion and process it into conclusions about the public's partiality towards a candidate. Based on these problems, this study provides a solution using a sentiment analysis approach to the public in favor of the Indonesian Presidential candidate based on Twitter data. Because public opinion in Indonesia has the potential to be valuable information used in the object of sentiment analysis [15]. Through the sentiment analysis approach, public opinion can be used as a reference in decision making, especially in the political contestation of the Presidential election. This research uses the Naïve Bayes classification method with TF-IDF weighting techniques and utilizes the Twitter API to obtain the latest data from figures proposed to be Indonesian Presidential candidates.

2. Research Method

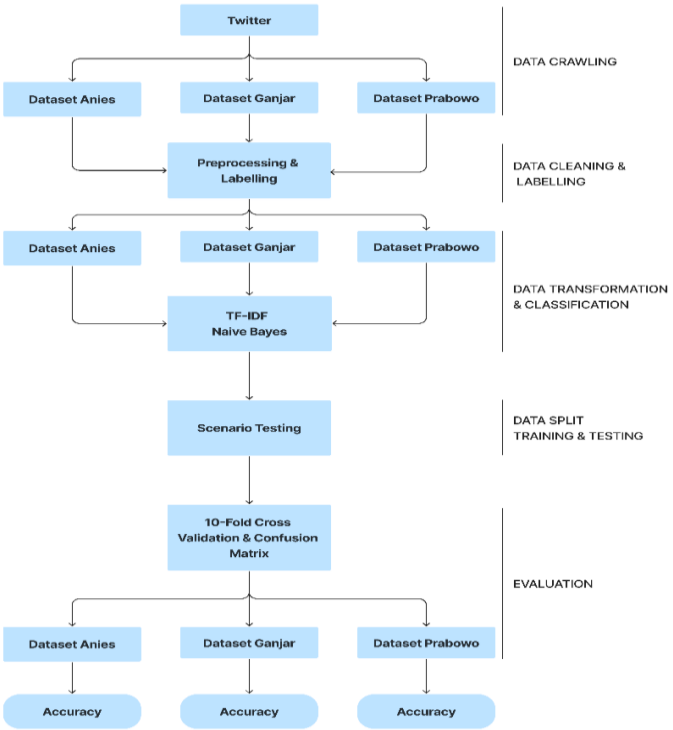


Figure 1. Flowchart Sentiment Analysis of Presidential Candidate

Sentiment analysis is one area of text mining science. In business, sentiment analysis is often used to analyze a company's products and services. This information is then processed to be used in decision-making [14]. Sentiment analysis uses text analysis, Natural Language Processing (NLP), and computational techniques to automate the extraction or classification of sentiment into sentiment rankings. Main purpose of sentiment analysis is to analyze rankings and calculate opinion scores. The score obtained can be divided into positive, negative, or neutral scores called polarity [17][18].

Figure 1 showing the study stage from the data crawling phase to the evaluation phase. The system consists of four main parts covering the processes of data collection, feature extraction, data labeling, and data classification using deep learning [19]. This study used three datasets, namely the Anies Baswedan, Ganjar Pranowo, and Prabowo Subianto datasets. Each dataset will go through steps of preprocessing and labeling data. The data labeling stage uses the scikit-learn library in python. Each result of each dataset will be reprocessed through the weighting stages with the TF-IDF and classifying the naïve Bayes method. Each dataset uses a test scenario to split the classification dataset into test and training data. The best results from the test scenario will be evaluated using 10-fold cross-validation and confusion matrix. Results of this stage, each dataset will provide the results of their respective accuracy.

2.1 Crawling Data

The data collected uses crawling techniques against tweets on the social media platform Twitter. This data crawling process uses jupyter notebook and anaconda tools with a programming language, namely python 3. The data taken is public opinion or tweets on the topic of 2024 presidential election.

2.2 Preprocessing

Preprocessing removes existing characters and words that do not match the document. Preprocessing stage makes text processing easier and better [20]. The data that has been generated will first be through preprocessing, which aims to change data that is not structured, is still rough and has noise so that it turns into a form of data that is ready to be processed [21]. The data generated through the crawling process from Twitter is in the form of data that is not structured, has no meaning, and even contains noise. Improvement of data in the form of preprocessing process text is processed so it can be used in the classification process [16]. The stages in preprocessing are case folding, tokenizing, stopword removal, normalization, and stemming, as shown in Figure 2.

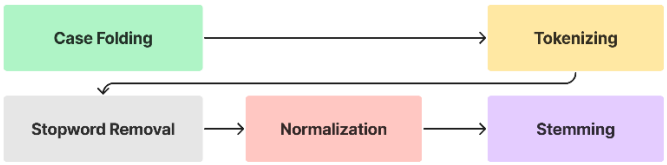


Figure 2. Data Preprocessing Stage

Figure 2 shows the five stages of the pretreatment process. The stage of case folding is to change the text to lowercase so that it is easy to classify. Tokenizing process i.e., the text will be cleared of symbols and other unused noise. If tokenization removes symbols with no values and noise, the stop word removal phase removes words with no values, such as conjunctions "yg", and "dgn". In the text normalization phase, a normalization process is performed to transform the text in the word stemming phase into basic word forms, facilitating the weighting and classification process.

2.3 Labelling

This phase is labeled using the scikit-learn library available in Python. This library is used to perform the analysis. Scikit-learn is a library for modeling and predicting new case instances [22][23].

2.4 Term Weighting

Term weighting is a technique for weighting documents. Several aspects affect the weighting, including term frequency (TF), inverse document frequency (IDF), and normalization. The weighting technique used in this study was to use the TF-IDF obtained from Equation 1 [24][25].

$$TF.IDF = TF_{i,j}\ x\ IDF_{i,j} = TF_{i,j}\ x\ log\frac{N}{DF_j} \quad (1)$$

Equation 1 shows the calculation of the occurrence of term j in document i or also called the frequency term denoted by $TF_{i,j}$. As well as calculating the frequency of occurrence of the entire document collection is also called the

frequency of occurrence of documents denoted by $DF_{i,j}$ [25]. Where N is the Collection document number, TF is Term Frequency, and IDF is Inverse Document Frequency.

2.5 Naïve Bayes

Naive Bayes is the basic concept of Bayes' Theorem, a data mining technique for classifying data. A dataset or classification process can be divided into her two phases: learning/training and testing/classification to obtain suitable classification parameters [26][27]. The Naive Bayes Classifier is a method for classifying derivatives. Bayes' theorem can also be used to maximize the posterior probability and increase the probability of classification into variables and conditional factors [28]. Bayesian methods are used to calculate the probability of an event based on observations made. A simple probabilistic Bayesian naive method makes predictions based on the class membership assumptions obtained from the Equation 2 [29].

$$P(H|X) = \frac{(P(X|H).P(H)}{P(X)} \quad (2)$$

Equation 2 shows the calculation of three probability values, namely the initial probability of each class P(H) from the training data, the conditional probability of each attribute distribution P(X|H) and the initial probability of each class with P(H) * P(X|H) [30]. Where X is the unknown class of data, H is the X data (specific class), P(H|X) is the probability of hypothesis H given condition X (posterior probability), P(H) is the H probability (Probability Prior), P(X|H) is the probability X based on the hypothesis condition H, and P(X) is Probability X.

2.6 K-Fold Cross Validation and Confusion Matrix

Evaluation of film assessment details is done according to the k-fold cross-validation method with value k = 10. This evaluation method is widely used in part for text classification [31]. Cross-validation (K-Fold) is a method used to evaluate predictions divided into training and test samples. Most data partitions will be divided to train the model, and the other small part will be used for testing. After that, it will be repeated for a certain time so that it can determine the errors that occur each time [32]. This study was tested with a test scenario, as shown in Table 1. The test process used is 10-fold cross-validation which will be iterated ten times in each test scenario, as shown in Figure 3.

Table 1. Split Scenarios for Testing

Training (%)	Testing (%)
70	30
80	20
90	10

Table 1 shows that data sharing is done based on the model used. Three scenarios are used for data exchange. Best case scenario to be used in the model evaluation calculation.

Figure 3. 10-Fold Cross Validation

Figure 3 describes the k-fold cross-validation test used as in Figure 3. It can be seen that the test data used follows the number of repetitions that occur. The best precision produced on iterations is the precision used as precision in cross-validation tests. Get confusion matrix from Equation 3 [33].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

Where TP (true positives) is the set of positive data classified as true by the system, TN (true negatives) is the set of negative data classified as true by the system, and FP (false positives) are classified as A set of positive data that has been FN (False Negative) is the amount of negative data incorrectly classified by the system.

3. Results and Discussion

The data used in this study comes from the social media platform Twitter. Data was obtained using crawling techniques in a Jupyter notebook. The retrieved data is keyword data that does not need to be defined in advance [34]. Acquired data is data for three months for each variable of 5000 tweets. The data that has been obtained is then saved into a file with a .xlsx or .csv extension.

The next stage is the preprocessing stage, which includes deduplication, case folding, tokenization, stopword removal, and stemming steps. A preprocessing phase is performed to ensure that the quality of the tweet data is more optimal during classification.

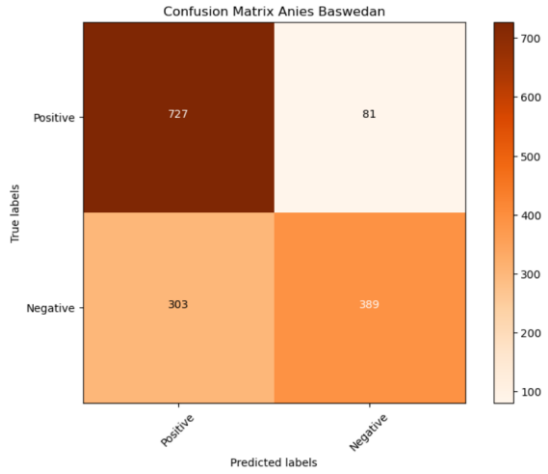
Data that passes the preprocessing phase is subject to sentiment labeling. This study used dual emotions: positive emotions and negative emotions. Excel or CSV files that have gone through the preprocessing and labeling process are weighted with TF-IDF. Based on the labeling stage using 5000 tweets on each candidate, Anies, Ganjar, and Prabowo, negative and positive sentiments were labeled as in Table 2.

Table 2. Results of Candidates Sentiment

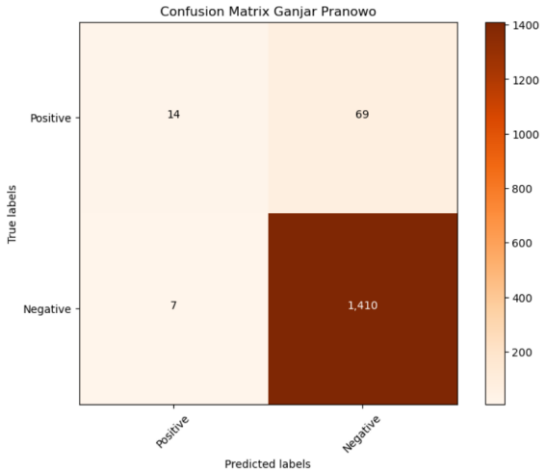
Candidate	Negative	Positive
Anies	2797	2203
Ganjar	542	4458
Prabowo	271	4279

Table 2 shows the results of labeling datasets into positive and negative sentiment classes. The data that has been obtained is then processed to the TF-IDF stage. It will then be determined polarity to use as a classification and test data model. The naïve Bayes classification method will form a data model with the division of test scenarios as in Table 1.

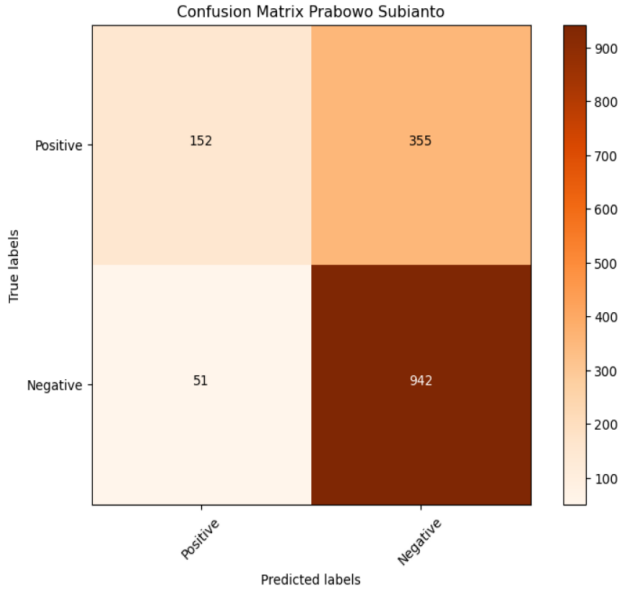
Determining the best model is done using test scenarios on each dataset. Best scenario results were obtained based on three test scenarios. That is, it split 70% training data and 30% testing data with up to 95% accuracy. After testing using a test scenario, the best test scenario is tested for validation against a classification system built using K-fold cross-validation with a value of K, which is 10. The datasets used are Anies, Ganjar, and Prabowo, with successive accuracy of 74%, 95%, and 73%. Based on testing using the confusion matrix shown in Figure 4(a), Figure 4(b), and Figure 4(c).



(a)



(b)



(c)

Figure 4. Confusion Matrix of Candidates

Figure 4 (a) shows the testing of the classification model using the confusion matrix for the dataset "Anies Presiden 2024," which got an accuracy result of 74,4%. Then figure 4 (b) shows the classification model testing using the confusion matrix for the dataset "Ganjar Presiden 2024," which got an accuracy result of 94.93%. And figure 4 (c) shows the classification model testing with the dataset "Prabowo Presiden 2024," which got an accuracy result of 72.93%.

The Comparison of accuracy the three datasets of Anies, Ganjar, and Prabowo using the K-Fold Cross Validation and Confusion Matrix is shown in Figure 5. Based on comparison data, Figure 5 shows the best results for k-fold cross-validation and confusion matrix on the Ganjar dataset. Meanwhile, the Anies dataset is superior in accuracy and confusion matrix validation test compared to the Prabowo dataset.

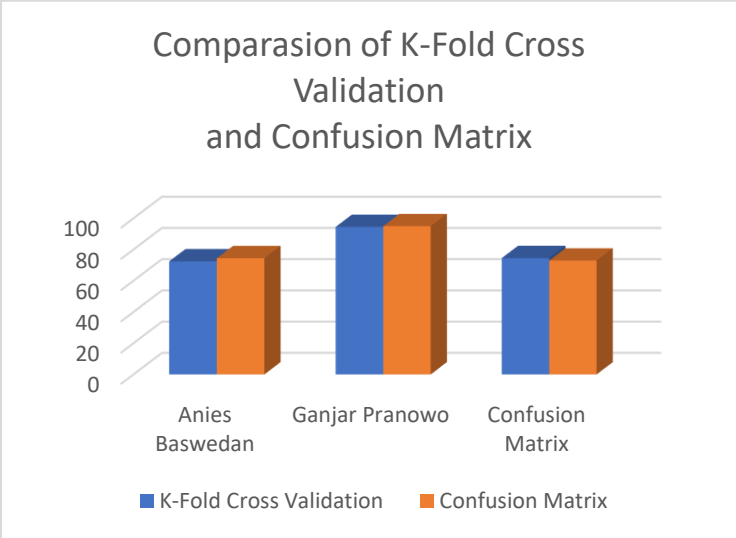


Figure 5. Comparison of K-Fold CV and Confusion Matrix Candidates

The f1-score in the positive and negative sentiment classes for each candidate was obtained, as shown in Figure 6. The highest negative sentiment class f-1 score results were obtained in the Anies dataset and the lowest in the Ganjar dataset. Meanwhile, in the positive sentiment class, the highest score was obtained in the Ganjar dataset and the weakest in the Anies dataset.

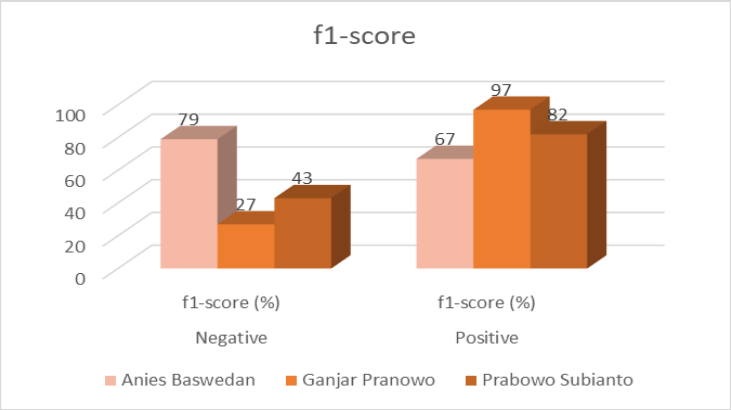


Figure 6. Candidates F1-Score Results

Based on the results of this study which has an approach in the equation of data sources, classification methods and data processing techniques used in the research of V. A. Fitri, et al (2019) [14], a comparison can be made that the research discussing LGBT topics has three classes of sentiment and obtained an accuracy of 86.43%. While in this study using two classes of sentiment, the best accuracy results were obtained 94.93%. Furthermore, in the research of M. Qorib, et al (2023) [15] with the same dataset, weighting and classification method as the topic of Covid-19 obtained an accuracy of 80.78%. Another study [32] which discussed the assessment of passengers on their voyage on one of the Saudi Arabian ships with the same dataset, weighting techniques and classification methods obtained an accuracy of 90.08%. This proves that this study has a better level of accuracy than some of these studies.

4. Conclusion

This study used a Twitter dataset of 15000, with the sentiment classes used being positive and negative. Before classifying, the data obtained will be given weight using TF-IDF. This technique will provide better accuracy when performing classification. Tests are run to get the best model using test scenarios that share training and test data. The best-case scenario consisted of 70% of the training data and 30% of the test data experimented on each dataset. Validation tests are performed with k-fold cross-validation and confusion matrices using these scenarios. K-fold cross-validation was performed using 10 iterations for each dataset, and the accuracy obtained for each dataset was 72.4% for the Anies dataset, 94.46% for Ganjar, and 74.5% for Prabowo. The validation using the confusion matrix obtained accuracy results on the Anies dataset 74.4%, Ganjar 94.93%, and Prabowo 72.93%. However, in the f1-score, the lowest score obtained in the positive class is in the Anies dataset, while in the negative class, namely the Ganjar dataset.

Based on previous research that has been carried out [35] using a total of 533 tweet data consisting of Ganjar Pranowo datasets as many as 274, Anies Baswedan as many as 120, Prabowo Subianto as many as 72, and Ridwan Kamil as many as 67. This study got the best accuracy with the Ganjar dataset on the 7th K-Fold, 73.68%. The labeling stage in this study uses the TextBlob library with three classes of sentiment, namely positive, neutral, and negative. Based on this, the updates increase the number of datasets, using two classes of positive and negative sentiments and the scikit-learn library in labeling text.

It can be concluded that the number of datasets and labeling techniques can affect the level of accuracy in sentiment analysis problems. This is evidenced by comparisons between this study and previous studies using the same method. Based on the tests, the most positive sentiment was obtained among the three candidates, namely the Anies dataset of 833 tweets. Meanwhile, the most negative sentiment was obtained in the Ganjar dataset of 1423. This research provides recommendations in subsequent studies to solve the problem of removing duplicate data, which is considered one of the causes of the effectiveness of data to be classified. Based on these various exposures, the Naïve Bayes method is one of the effective methods that can be used in sentiment analysis cases because it provides a high level of accuracy.

References

[1] Admin, “Tahapan dan Jadwal Penyelenggaraan Pemilu Tahun 2024,” Komisi Pemilihan Umum, 2022.
[2] Admin, “Survei Poltracking: Ganjar, Prabowo, Anies Jadi Capres Terkuat di 2024,” Poltracking Indonesia, 2022.
[3] I. Riadi, D. Aprilliansyah, and S. Sunardi, “Mobile Device Security Evaluation using Reverse TCP Method,” *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, 2022, https://doi.org/10.22219/kinetik.v7i3.1433.
[4] A. Karami *et al.*, “2020 U.S. presidential election in swing states: Gender differences in Twitter conversations,” *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, 2022, https://doi.org/10.1016/j.jjimei.2022.100097.
[5] A. W. Pradana and M. Hayaty, “The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts,” *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, pp. 375–380, 2019, https://doi.org/10.22219/kinetik.v4i4.912.

[6] N. Meliana, Sunardi, and A. Fadlil, "Identification of Cyber Bullying by using Clustering Methods on Social Media Twitter," *J. Phys. Conf. Ser.*, vol. 1373, no. 1, 2019, https://doi.org/10.1088/1742-6596/1373/1/012040.
[7] I. Riadi, H. Herman, and A. Z. Ifani, "Optimization of System Authentication Services using Blockchain Technology," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, 2021, https://doi.org/10.22219/kinetik.v6i4.1325.
[8] H. Sujaini, "Performance of Methods in Identifying Similar Languages Based on String to Word Vector," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 6, no. 1, pp. 9–14, 2020, https://doi.org/10.23917/khif.v6i1.8199.
[9] A. Abayomi-Alli, O. Abayomi-Alli, S. Misra, and L. Fernandez-Sanz, "Study of the Yahoo-Yahoo Hash-Tag Tweets Using Sentiment Analysis and Opinion Mining Algorithms," *Inf.*, vol. 13, no. 3, pp. 1–22, 2022, https://doi.org/10.3390/info13030152.
[10] S. Nurul, J. Fitriyyah, N. Safriadi, E. Esyudha, and P. #3, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *(Jurnal Edukasi dan Penelit. Inform.*, vol. 5, no. 3, pp. 279–285, 2019, https://doi.org/10.26418/jp.v5i3.34368.
[11] A. M. Soesanto, C. Chandra, A. M. Soesanto, and C. Chandra, "ScienceDirect Sentiments comparison on on Twitter Twitter about about LGBT LGBT Sentiments comparison Sentiments comparison Twitter about LGBT," *Procedia Comput. Sci.*, vol. 216, pp. 765–773, 2023, https://doi.org/10.1016/j.procs.2022.12.194.
[12] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. R. I. M. Setiadi, and R. S. Basuki, "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 2, pp. 799–806, 2020, https://doi.org/10.12928/TELKOMNIKA.V18I2.14744.
[13] M. Liang and T. Niu, "Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs," *Procedia Comput. Sci.*, vol. 208, pp. 460–470, 2022, https://doi.org/10.1016/j.procs.2022.10.064.
[14] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, https://doi.org/10.1016/j.procs.2019.11.181.
[15] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst. Appl.*, vol. 212, no. September 2022, p. 118715, 2023, https://doi.org/10.1016/j.eswa.2022.118715.
[16] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, and S. H. Malik, "Detecting twitter hate speech in COVID-19 era using machine learning and ensemble learning techniques," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100120, 2022, https://doi.org/10.1016/j.jjimei.2022.100120.
[17] F. Alzami, E. D. Udayanti, D. P. Prabowo, and R. A. Megantara, "Document Preprocessing with TF-IDF to Improve the Polarity Classification Performance of Unstructured Sentiment Analysis," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, pp. 235–242, 2020, https://doi.org/10.22219/kinetik.v5i3.1066.
[18] V. P. Ramadhan, P. Purwanto, and F. Alzami, "Sentiment Analysis of Community Response Indonesia Against Covid-19 on Twitter Based on Negation Handling," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 2, 2022, https://doi.org/10.22219/kinetik.v7i2.1429.
[19] B. Kholifah, I. Syarif, and T. Badriyah, "Mental Disorder Detection via Social Media Mining using Deep Learning," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, pp. 309–316, 2020, https://doi.org/10.22219/kinetik.v5i4.1120.
[20] A. Yudhana, A. Fadlil, and M. Rosidin, "Indonesian words error detection system using nazief adriani stemmer algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12, pp. 219–225, 2019, https://doi.org/10.14569/ijacsa.2019.0101231.
[21] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 8, no. 1, pp. 59–71, 2022, https://doi.org/10.23917/khif.v8i1.15213.
[22] M. Shaden, A. Fadel, S. Achmad, and R. Sutoyo, "ScienceDirect ScienceDirect Sentiment analysis for customer review : Case study of Traveloka Sentiment analysis for customer review : Case study of Traveloka," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 682–690, 2023, https://doi.org/10.1016/j.procs.2022.12.184.
[23] D. Samuel, L. Aparecido, A. Adeel, and J. Paulo, "PL-kNN : A Python-based implementation of a parameterless ? -Nearest Neighbors classifier," *Softw. Impacts*, vol. 15, no. November 2022, p. 100459, 2023, https://doi.org/10.1016/j.simpa.2022.100459.
[24] I. Riadi, S. Sunardi, and P. Widiandana, "Mobile Forensics for Cyberbullying Detection using Term Frequency - Inverse Document Frequency (TF-IDF)," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 2, p. 68, 2020, https://doi.org/10.26555/jiteki.v5i2.14510.
[25] Imamah and F. H. Rachman, "Twitter sentiment analysis of Covid-19 using term weighting TF-IDF and logistic regresion," *Proceeding - 6th Inf. Technol. Int. Semin. ITIS 2020*, pp. 238–242, 2020, https://doi.org/10.1109/ITIS50118.2020.9320958.
[26] D. Jollyta, G. Gusrianty, and D. Sukrianto, "Analysis of Slow Moving Goods Classification Technique: Random Forest and Naïve Bayes," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 5, no. 2, pp. 134–139, 2019, https://doi.org/10.23917/khif.v5i2.8263.
[27] I. Riadi, R. Umar, and F. D. Aini, "Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm)," *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24, 2019, https://doi.org/10.33096/ilkom.v11i1.361.17-24.
[28] A. Yudhana, D. Sulistyo, and I. Mufandi, "GIS-based and Naïve Bayes for nitrogen soil mapping in Lendah, Indonesia," *Sens. Bio-Sensing Res.*, vol. 33, p. 100435, 2021, https://doi.org/10.1016/j.sbsr.2021.100435.
[29] A. Yudhana, I. Riadi, and F. Ridho, "DDoS classification using neural network and naïve bayes methods for network forensics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 11, pp. 177–183, 2018, https://doi.org/10.14569/ijacsa.2018.091125.
[30] N. Hayatin, G. I. Marthasari, and L. Nuraini, "Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization," *J. Online Inform.*, vol. 5, no. 1, pp. 81–88, 2020, https://doi.org/10.15575/join.v5i1.558.
[31] F. F. Zain and Y. Sibaroni, "Effectiveness of SVM Method by Naïve Bayes Weighting in Movie Review Classification," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 5, no. 2, pp. 108–114, 2019, https://doi.org/10.23917/khif.v5i2.7770.
[32] B. Al sari *et al.*, "Sentiment analysis for cruises in Saudi Arabia on social media platforms using machine learning algorithms," *J. Big Data*, vol. 9, no. 1, 2022, https://doi.org/10.1186/s40537-022-00568-5.
[33] F. Rahmad, Y. Suryanto, and K. Ramli, "Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 879, no. 1, pp. 1–12, 2020, https://doi.org/10.1088/1757-899X/879/1/012076.
[34] K. Brito and P. J. L. Adeodato, "Machine learning for predicting elections in Latin America based on social media engagement and polls," *Gov. Inf. Q.*, vol. 40, no. 1, 2023, https://doi.org/10.1016/j.giq.2022.101782.
[35] M. Raihan, F. Sya' Bani |, F. Sya' Bani, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes," *J. Ris. Komputer)*, vol. 9, no. 2, pp. 2407–389, 2022, https://doi.org/10.30865/jurikom.v9i2.3989.